

**TWO-STEP SOURCE SIDE IMPLANT FOR IMPROVING SOURCE  
RESISTANCE AND SHORT CHANNEL EFFECT IN DEEP SUB-0.18 $\mu$ m FLASH  
MEMORY TECHNOLOGY**

5

Zhigang Wang

Yue-Song He

Richard Fastow

**BACKGROUND OF THE INVENTION**

10     **1.     Field of the Invention**

          This invention relates generally to the art of microelectronic integrated circuits. More specifically, this invention relates to high performance microelectronic flash memory cells and to the art of manufacturing high performance microelectronic flash memory cells. Even more specifically, this invention relates to high performance  
15     microelectronic flash memory cells with reduced source resistance and improved short channel effect in deep sub-0.18 $\mu$ m flash memory technology.

**2.     Discussion of the Related Art**

          A microelectronic flash or block erase Electrically Erasable Programmable Read-Only Memory (Flash EEPROM) includes an array of cells that can be independently  
20     programmed and read. The size of each cell and thereby the memory array are made small by omitting transistors known as select transistors that enable the cells to be erased independently. As a result, all of the cells are erased together as a block.

          A memory of this type includes individual Metal-Oxide-Semiconductor (MOS) field effect transistor memory cells, each of which includes a source, a drain, a floating  
25     gate and a control gate to which various voltages are applied to program the cell with a binary 1 or 0, or to erase all of the cells as a block.

          The cells are connected in an array of rows and columns, with the control gates of the cells in a row being connected to a respective wordline and the drains of the cells in a column being connected to a respective bitline. The sources of the cells in either a column

or a row are connected together and each column or row common source connections are then connected to a common source voltage  $V_{SS}$ . This arrangement is known as a NOR flash memory configuration.

5 A cell is programmed by applying a voltage, typically 9 volts to the control gate, applying a voltage of approximately 5 volts to the drain and grounding the common voltage source  $V_{SS}$ , which causes hot electrons to be injected from a drain depletion region into the floating gate. Upon removal of the programming voltages, the injected electrons are trapped in the floating gate and create a negative charge therein which increases the threshold voltage of the cell to a value in excess of approximately 4 volts.

10 A cell is read by applying typically 5 volts to the control gate, applying 1 volt to the bitline to which the drain is connected, grounding the common source voltage  $V_{SS}$ , and sensing the bitline current. If the cell is programmed and the threshold voltage is relatively high (4 volts), the bitline current will be zero or at least relatively low. If the cell is not programmed or erased, the threshold voltage will be relatively low (2 volts), the control gate voltage will enhance the channel, and the bitline current will be relatively high.

20 A cell can be erased in several ways. In one arrangement, a cell is erased by applying a relatively high voltage, typically 12 volts, to the source, grounding the control gate and allowing the drain to float. This causes the electrons that were injected into the floating gate during programming to undergo Fowler-Nordheim tunneling from the floating gate through the thin tunnel oxide layer to the source. A cell can also be erased by applying a negative voltage on the order of minus 10 volts to the control gate, applying 5 volts to the source and allowing the drain to float. Another method of erasing is by applying 5V to the P-well and minus 10V to the control gate while allowing the source/drain to float.

30 However, as the dimensions of the flash memory array have been aggressively scaled down and the product arrays produced with ultra high density, the greatest challenge for deep sub-0.18 $\mu$ m high performance non-volatile memory cell design is to control the short channel dimension in order to control the short channel effects, such as  $V_t$  rolloff, high DIBL and excess column leakage, accompanied with less tolerance of the polysilicon gate length variation across the product array.

Therefore, what is needed is a method of providing a lightly doped source junction near the critical gate region and a heavily doped source junction away from the critical gate region in such a way that the lateral diffusion is decreased while maintaining a low  $V_{SS}$  resistance.

5     **SUMMARY OF THE INVENTION**

According to the present invention, the foregoing and other objects and advantages are obtained by a method of manufacturing flash memory devices having lightly doped source regions near the critical gate region and heavily doped source regions away from the critical gate region.

10     In accordance with one aspect of the invention, in a first embodiment a first source mask is formed exposing source regions and portions of the gates and n dopant ions are implanted in the exposed regions. The first source mask is removed and a second source mask is formed exposing a portion of the source regions and  $n^+$  dopant ions are implanted in the exposed regions.

15     In accordance with another aspect of the invention, in a second embodiment a source mask is formed exposing a portion of the source regions and  $n^+$  ions are implanted in the exposed regions.

The described method thus controls short channel effects such as  $V_t$  rolloff, high DIBL and excess column leakage with less gate length variation across the product array.

20     The present invention is better understood upon consideration of the detailed description below, in conjunction with the accompanying drawings. As will become readily apparent to those skilled in the art from the following description, there is shown and described embodiments of this invention simply by way of illustration of the best mode to carry out the invention. As will be realized, the invention is capable of other  
25     embodiments and its several details are capable of modifications in various obvious aspects, all without departing from the scope of the invention. Accordingly, the drawings and detailed description will be regarded as illustrative in nature and not as restrictive.

**BRIEF DESCRIPTION OF THE DRAWINGS**

The novel features believed characteristic of the invention are set forth in the appended claims. The invention itself, however, as well as a preferred mode of use, and further objects and advantages thereof, will best be understood by reference to the following detailed description of an illustrative embodiment when read in conjunction with the accompanying drawings, wherein:

**Figures 1A – 1D** show a prior art source side implant process;

**Figure 1A** is a top view of a portion of a flash memory device showing the position of a first source mask formed on the portion of the flash memory device;

**Figure 1B** is a cross sectional view of the portion of the flash memory device shown in **Figure 1A** being implanted with  $n^+$  ions;

**Figure 1C** is a cross sectional view of the portion of the flash memory device as shown in **Figure 1B** with the first source mask removed, a second source mask formed and being implanted with  $n$  ions;

**Figure 1D** is the portion of the flash memory device as shown in **Figure 1C** with the second source mask removed and after an anneal process that has driven the  $n$  ions into the substrate forming a final dopant profile;

**Figures 2A – 2E** show a first embodiment of a source side implant process in accordance with the present invention;

**Figure 2A** is a top view of a portion of a flash memory device showing the position of a first source mask formed on the portion of the flash memory device;

**Figure 2B** is cross sectional view of the portion of the flash memory device as shown in **Figure 2A** being implanted with  $n$  ions;

**Figure 2C** is a top view of the portion of the flash memory device as shown in **Figure 2A** showing the position of a second source mask formed on the portion of the flash memory device;

**Figure 2D** is a cross sectional view of the portion of the flash memory device as shown in **Figure 2C** being implanted with  $n^+$  ions;

**Figure 2E** shows the portion of the flash memory device as shown in **Figure 2D** with the second source mask removed and after an anneal process that has driven the  $n$  ions into the substrate forming a desired dopant profile;

**Figures 3A – 3C** show a second embodiment of a source side implant process in accordance with the present invention;

**Figure 3A** is a top view of a portion of a flash memory device showing the position of a source mask formed on the portion of the flash memory device;

**Figure 3B** is a cross sectional view of the portion of the flash memory device as shown in **Figure 3A** being implanted with  $n^+$  ions; and

**Figure 3C** shows the portion of the flash memory device as shown in **Figure 3B** with the source mask removed and after an anneal process that has driven the  $n^+$  ions into the substrate forming a desired dopant profile.

#### **DETAILED DESCRIPTION**

Reference is now made in detail to a specific embodiment or specific embodiments of the present invention that illustrate the best mode or modes presently contemplated by the inventors for practicing the invention.

**Figures 1A – 1D** illustrate selected steps of a prior art source side implant process.

**Figure 1A** is a top view of a portion **100** of a semiconductor flash memory. The portion **100** of the semiconductor flash memory shows the drain regions **102**, the gates **104**, and the shared source region **106** of a first pair of transistors **107**, the drain regions **108**, the gates **110**, and the shared source region **112** of a second pair of transistors **113**, and the drain regions **114**, the gates **116**, and the shared source region **118** of a third pair of transistors **119**. The shaded portion indicates a first source mask **120** formed on the portion **100** of the semiconductor flash memory. The lines (Stacked Gate Edge) **SGE 122** indicate the edges of the gates **104**, **110** and **116**. The edge of the shaded portion of the first source mask **120** coincides with the edges of the **SGE 122** adjacent the shared source regions **106**, **112** and **118**.

**Figure 1B** is a cross sectional view of the portion **100** of the semiconductor memory as shown in **Figure 1A** showing the drain regions **114**, the gates **116** and the shared source **118** of the third pair of transistors **119**. The portion **100** of the semiconductor memory is shown being implanted with  $n^+$  ions indicated by the arrows

5 **124**. The dotted line **126** indicates the initial profile of the  $n^+$  ions as implanted.

**Figure 1C** is the cross sectional view of the portion **100** of the flash memory device as shown in **Figure 1B** with the first source mask **120** (**Figures 1A & 1B**) removed and a second source mask **128** formed on the portion **100** of the flash memory device. The portion **100** of the semiconductor memory is shown being implanted with  $n$  ions indicated

10 by the arrows **130**. The dotted line **132** indicates the initial dopant profile of the  $n$  ions as implanted. As is known in the semiconductor manufacturing art, an  $n^+$  or an  $n$  indicates the concentration and energy of the implantation of ions into a semiconductor device. The  $n^+$  indicates a heavy or high dosage and is typically an implantation of arsenic ions. The  $n$  indicates a relatively light dosage and is typically phosphorus ion, however the light

15 dosage could also be arsenic ions.

**Figure 1D** is portion **100** of the flash memory device as shown in **Figure 1C** with the second source mask **128** (**Figure 1C**) removed and after an anneal process that has driven the  $n^+$  and  $n$  ions into the substrate to form the final dopant profile. It should be noted that the  $n^+$  and  $n$  ions extend beyond the edge of the gate **116** toward the drain

20 regions **114**. Because of the extension of the  $n^+$  and  $n$  profile underneath the gates **116** the channel formed underneath the gates **116** is shortened.

**Figures 2A – 2E** show a first embodiment of a source side implant process in accordance with the present invention. **Figure 2A** is a top view of a portion **200** of a semiconductor flash memory. The portion **200** of the semiconductor flash memory shows

25 the drain regions **202**, the gates **204**, and the shared source region **206** of a first pair of transistors **207**, the drain regions **208**, the gates **210**, and the shared source region **212** of a second pair of transistors **213**, the drain regions **214**, the gates **216**, and the shared source region **218** of a third pair of transistors **219**. The shaded portions indicate a first source mask **220** formed on the portion **200** of the semiconductor flash memory. The lines

30 (Stacked Gate Edge) SGE **222** indicate the edges of the gates **204**, **210** and **216**. The edge **223** of the shaded portion of the first source mask **220**, unlike the prior art process shown

in **Figure 1A**, does not coincide with the edges **225** of the SGE **222** adjacent the shared source regions **206**, **212** and **218**.

**Figure 2B** is a cross sectional view of the portion **200** of the semiconductor memory as shown in **Figure 2A** showing the drain regions **214**, the gates **216** and the shared source **218** of the third pair of transistors **219**. The portion **200** of the semiconductor memory is shown being implanted with n ions indicated by the arrows **224**. The dotted line **226** indicates the initial profile of the n ions as implanted. The edges **223** of the first source mask are shown not coinciding with the edges **225** of the gates **216**.

**Figure 2C** is a top view of a portion **200** of a semiconductor flash memory as shown in **Figure 2A** with the first source mask **220** removed and a second source mask **228** formed on the portion **200** of the semiconductor flash memory. However, unlike the first source mask **220** in **Figure 2A**, the edges **223** of the second source mask **228** extend beyond the edges **225** of the SGE **222**.

**Figure 2D** is a cross sectional view of the portion **200** of the flash memory device as shown in **Figure 2C**. The portion **200** of the semiconductor memory is shown being implanted with  $n^+$  ions indicated by the arrows **230**. The dotted line **232** indicates the initial dopant profile of the  $n^+$  ions as implanted. The edges **223** of the second mask are shown not coinciding with the edges **225** of the gates **216**.

**Figure 2E** is portion **200** of the flash memory device as shown in **Figure 2D** with the second source mask **228** (**Figure 2D**) removed and after an anneal process that has driven the  $n^+$  and n ions into the substrate. It should be noted that the  $n^+$  and n ions do not extend beyond the edges **225** of the gates **216** toward the drain regions **214**. This allows the overall dimensions of the gate to be reduced because the channel length is not shortened underneath the gates **216**.

**Figures 3A – 3C** show a second embodiment of a source side implant process in accordance with the present invention. **Figure 3A** is a top view of a portion **300** of a semiconductor flash memory. The portion **300** of the semiconductor flash memory shows the drain regions **302**, the gates **304**, and the shared source region **306** of a first pair of transistors **307**, the drain regions **308**, the gates **310**, and the shared source region **312** of a second pair of transistors **313**, the drain regions **314**, the gates **316**, and the shared source region **318** of a third pair of transistors **319**. The shaded portions indicate a source mask

320 formed on the portion 300 of the semiconductor flash memory. The lines (Stacked Gate Edge) SGE 322 indicate the edges of the gates 304, 310 and 316. The edge 323 of the shaded portion of the first source mask 320, unlike the prior art process shown in Figure 1A, does not coincide with the edges 325 of the SGE 322 adjacent the shared source regions 306, 312 and 318.

Figure 3B is a cross sectional view of the portion 300 of the semiconductor memory as shown in Figure 3A showing the drain regions 314, the gates 316 and the shared source 318 of the third pair of transistors 319. The portion 300 of the semiconductor memory is shown being implanted with  $n^+$  ions indicated by the arrows 324. The dotted line 326 indicates the initial profile of the  $n^+$  ions as implanted. The edges 323 of the source mask are shown not coinciding with the edges 325 of the gates 316.

Figure 3C is portion 300 of the flash memory device as shown in Figure 3B with the source mask 320 (Figure 3B) removed and after an anneal process that has driven the  $n^+$  ions into the substrate. It should be noted that the  $n^+$  ions do not extend beyond the edges 325 of the gates 316 toward the drain regions 314. This allows the overall dimensions of the gate to be reduced because the channel length is not shortened underneath the gates 316.

In summary, the described method controls short channel effects such as  $V_t$  rolloff, high DIBL and excess column leakage with less gate length variation across the product array.

The foregoing description of the embodiments of the invention has been presented for purposes of illustration and description. It is not intended to be exhaustive or to limit the invention to the precise form disclosed. Obvious modifications or variations are possible in light of the above teachings. The embodiments were chosen and described to provide the best illustration of the principles of the invention and its practical application to thereby enable one of ordinary skill in the art to utilize the invention in various embodiments and with various modifications as are suited to the particular use contemplated. All such modifications and variations are within the scope of the invention as determined by the appended claims when interpreted in accordance with the breadth to which they are fairly, legally, and equitably entitled.